Preimage Problems for Reaction Systems*

Alberto Dennunzio[1], Enrico Formenti[2], Luca Manzoni[1], and Antonio E. Porreca[1]

[1] Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca, Viale Sarca 336/14, 20126 Milano, Italy
{dennunzio,luca.manzoni,porreca}@disco.unimib.it
[2] Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
formenti@unice.fr

Abstract. We investigate the computational complexity of some problems related to preimages and ancestors of states of reaction systems. In particular, we prove that finding a minimum-cardinality preimage or ancestor, computing their size, or counting them are all intractable problems, with complexity ranging from $\mathbf{FP}^{\mathbf{NP}[\log n]}$ to **FPSPACE**(poly).

Keywords: reaction systems, computational complexity

1 Introduction

Recently many new computational models have been introduced. Most of them are inspired by natural phenomena. This is also the case of Reaction Systems (RS), proposed by Ehrenfeucht and Rozenberg in [2], which are a metaphor for basic chemical reactions. Informally, a reaction system is made of a (finite) set of entities (molecules) and a (finite) set of admissible reactions. Each reaction is a triple of sets: *reactants*, *inhibitors* and *products* (clearly the set of reactants and the one of inhibitors are disjoint). Given a set of reactants T, a reaction (R, I, P) is applied if $R \subseteq T$ and if there are no inhibitors (*i.e.* $T \cap I$ is empty); the result is the replacement of T by the set of products P. Given a set of reactants T, all admissible reactions are applied in parallel. The final set of products is the union of all single sets of products of each reaction which is admissible for T.

Studying RS is interesting for a number of reasons, not only as a clean computational model allowing precise formal analysis but also as a reference *w.r.t.* other computing systems. For example, in [5], the authors showed an embedding of RS into Boolean automata networks (BAN), a well-known model used in a number of application domains. Remark that for BAN the precise complexity of only for a bunch of problems about the dynamical behaviour is known. Via

* This work has been supported by Fondo d'Ateneo (FA) 2013 of Università degli Studi di Milano-Bicocca: "Complessità computazionale in modelli di calcolo bioispirati: Sistemi a membrane e sistemi a reazioni", by the Italian MIUR PRIN 2010-2011 grant "Automata and Formal Languages: Mathematical and Applicative Aspects" H41J12000190001, and by the French National Research Agency project EMC (ANR-09-BLAN-0164).

the embedding of RS into BAN, all the complexity results about RS are indeed lower bounds for the corresponding ones for BAN.

In this paper, we continue the exploration of the computational complexity of properties of RS. The focus is on preimages and ancestors of minimal size. In more practical terms, this could be useful when minimising the number of chemical entities necessary to obtain a target compound. Indeed, given a current state T, the minimal pre-image (resp., n-th-ancestor) problem or MPP (resp., MAP) consists in finding the minimal set (*w.r.t.* cardinality) of reactants which produces T in one step (resp., n steps). Variants of MPP and MAP consider counting the pre-images (#MPP); counting the ancestors (#MAP); or computing the size of the minimal pre-image (SMPP) or of the minimal ancestor (SMAP). We prove that (see Section 2 for the precise definition of the complexity classes):

- MPP $\in \mathbf{FP}^{\mathbf{NP}}$ and it is $\mathbf{FP}_{\|}^{\mathbf{NP}}$-hard under metric reductions;
- #MPP is in $\#\mathbf{P}^{\mathbf{NP}[\log n]}$ and it is $\#\mathbf{P}$-hard under parsimonious reductions;
- SMPP is $\mathbf{FP}^{\mathbf{NP}[\log n]}$-complete under metric reductions;
- MAP and #MAP are complete for **FPSPACE**(poly) under metric reductions;
- SMAP is **FPSPACE**(log)-complete under metric reductions.

These results are important for further understanding the computational capabilities of RS but they also provide clean new items to the (relatively) short list of examples of problems in high functional complexity classes. Finally, remark that the problem of pre-image existence has been proved to be in **NP** by Salomaa [10]. However, here the complexity is higher because of the minimality requirement.

2 Basic Notions

We briefly recall the basic notions about RS [3]. In this paper we require the sets of reactants and inhibitors of a reaction to be nonempty, as is sometimes enforced in the literature; our results also hold when empty sets are allowed.

Definition 1. *Consider a finite set S, whose elements are called* entities*. A* reaction *a over S is a triple (R_a, I_a, P_a) of nonempty subsets of S. The set R_a is the set of* reactants*, I_a the set of* inhibitors*, and P_a is the set of* products*. The set of all reactions over S is denoted by* $\mathrm{rac}(S)$*.*

Definition 2. *A* Reaction System (RS) *is a pair $\mathcal{A} = (S, A)$ where S is a finite set, called the* background set*, and $A \subseteq \mathrm{rac}(S)$.*

Given a *state* $T \subseteq S$, a reaction a is said to be *enabled* in T when $R_a \subseteq T$ and $I_a \cap T = \varnothing$. The *result function* $\mathrm{res}_a \colon 2^S \to 2^S$ of a, where 2^S denotes the power set of S, is defined as $\mathrm{res}_a(T) = P_a$ if a is enabled in T, and $\mathrm{res}_a(T) = \varnothing$ otherwise. The definition of res_a naturally extends to sets of reactions: given $T \subseteq S$ and $A \subseteq \mathrm{rac}(S)$, define $\mathrm{res}_A(T) = \bigcup_{a \in A} \mathrm{res}_a(T)$. The result function $\mathrm{res}_{\mathcal{A}}$ of a RS $\mathcal{A} = (S, A)$ is res_A, i.e., it is the result function on the whole set of reactions.

Definition 3. *Let $\mathcal{A} = (S, A)$ be a RS. For any $T \subseteq S$, an element $U \subseteq S$ is an* ancestor *of T if $\mathrm{res}^t_{\mathcal{A}}(U) = T$ for some $t \in \mathbb{N}$. If $t = 1$, U is called* preimage *of T. An ancestor (resp., preimage) U of T is* minimal *if $|U| \leq |V|$ for all ancestors (resp., preimages) V of T.*

A state always admits at least itself as ancestor but might not have a preimage.

We describe the complexity of preimage and ancestor problems for RS with complexity classes of functions problems (see [8, 7] for further details). Let Σ be an alphabet. The class **FP** (resp., $\mathbf{FP}^{\mathbf{NP}}$) consists of all binary relations R over $\Sigma^\star$ having a "choice function" $f \subseteq R$ with $\mathrm{dom}\, f = \mathrm{dom}\, R$ that can be computed in polynomial time by a deterministic Turing machine (TM) without access to oracles (resp., with access to an oracle for an **NP** decision problem). Additional requirements on the oracle queries define the subclasses $\mathbf{FP}^{\mathbf{NP}[\log n]}$ and $\mathbf{FP}^{\mathbf{NP}}_{\parallel}$ of $\mathbf{FP}^{\mathbf{NP}}$, where the number of allowed oracle queries is $O(\log n)$ and the queries are performed in parallel (i.e., every current query string does not depend on the results of previous queries), respectively. We have $\mathbf{FP} \subseteq \mathbf{FP}^{\mathbf{NP}[\log n]} \subseteq \mathbf{FP}^{\mathbf{NP}}_{\parallel} \subseteq \mathbf{FP}^{\mathbf{NP}}$. The class **#P** consists of all functions $f\colon \Sigma^\star \to \mathbb{N}$ with a polynomial-time nondeterministic TM having exactly $f(x)$ accepting computations on every input x. If in addition $O(\log n)$ queries to an **NP** oracle are allowed, the class $\#\mathbf{P}^{\mathbf{NP}[\log n]}$ is defined. Clearly, $\mathbf{FP} \subseteq \#\mathbf{P} \subseteq \#\mathbf{P}^{\mathbf{NP}[\log n]}$. In this paper we will also refer to **FPSPACE**, i.e., the collection of binary relations having a choice function computable in polynomial space, and its two subclasses **FPSPACE**(poly) and **FPSPACE**(log), in which the output is limited to polynomial length and logarithmic length, respectively, rather than exponential length. Remark that **FPSPACE**(poly) is just $\natural$**PSPACE**, i.e., the class of functions $f\colon \Sigma^\star \to \mathbb{N}$ such that there exists a polynomial-space nondeterministic TM performing only a polynomial number of nondeterministic choices and having exactly $f(x)$ accepting computations on every input x.

Hardness for these classes is defined in terms of two kinds of reductions. The first one is the many-one reduction, also called *parsimonious* when dealing with counting problems: a function f is many-one reducible to g if there exists a function $h \in \mathbf{FP}$ such that $f(x) = g(h(x))$ for every input x. A generalisation is the *metric reduction* [6]: a function f is metric reducible to g if there exist functions $h_1, h_2 \in \mathbf{FP}$ such that $f(x) = h_2(x, g(h_1(x)))$ for every input x. These notions of reduction can be generalised to reductions between binary relations.

Since we do not deal with sublinear space complexity, without loss of generality, throughout this paper we assume that all TM computing functions use a unique tape, both for input and work. We also assume that they move their tape head to the leftmost cell before entering a final state.

3 Preimage Problems

First of all, inspired by the algorithm described in [9], we show a tight relation between the MPP and the problem of finding a minimal unary travelling salesman tour (TSP) [8], where the edge weights are encoded in unary and, hence, bounded by a polynomial in the number of vertices.

Lemma 4. *Finding a minimal unary TSP tour is metric reducible to* MPP.

Proof. Suppose we are given any set of vertices V, with $|V| = n$, and any unary-encoded weight function $w\colon V^2 \to \mathbb{N}$. We build a RS $\mathcal{A} = (S, A)$ admitting a state whose preimages encode the weighted simple cycles over V. The background set is given by $S = E \cup W \cup \{\heartsuit, \spadesuit\}$, where

$$E = \{(u,v)_t : u, v \in V \text{ and } 0 \le t < n\}$$
$$W = \{\diamondsuit_{(u,v),i} : u, v \in V \text{ and } 1 \le i \le w(u,v)\}$$

while the reactions in A, with u, v, u_1, u_2, v_1, v_2 ranging over V and t, t_1, t_2 ranging over $\{0, \dots, n-1\}$, are

$$(\{(u_1,v_1)_t, (u_2,v_2)_t, \heartsuit\}, \{\spadesuit\}, \{\spadesuit\}) \quad \text{if } (u_1,v_1) \neq (u_2,v_2) \quad (1)$$
$$(\{(u,v_1)_{t_1}, (u,v_2)_{t_2}, \heartsuit\}, \{\spadesuit\}, \{\spadesuit\}) \quad \text{if } v_1 \neq v_2 \text{ or } t_1 \neq t_2 \quad (2)$$
$$(\{(u_1,v)_{t_1}, (u_2,v)_{t_2}, \heartsuit\}, \{\spadesuit\}, \{\spadesuit\}) \quad \text{if } u_1 \neq u_2 \text{ or } t_1 \neq t_2 \quad (3)$$
$$(\{(u_1,v_1)_t, (u_2,v_2)_{(t+1) \bmod n}, \heartsuit\}, \{\spadesuit\}, \{\spadesuit\}) \quad \text{if } v_1 \neq u_2 \quad (4)$$
$$(\{\heartsuit\}, \{(u,v)_t : u, v \in V\} \cup \{\spadesuit\}, \{\spadesuit\}) \quad (5)$$
$$(\{(u,v)_t, \heartsuit\}, \{\diamondsuit_{(u,v),i}, \spadesuit\}, \{\spadesuit\}) \quad \text{for } 1 \le i \le w(u,v) \quad (6)$$
$$(\{\diamondsuit_{(u,v),i}, \heartsuit\}, \{(u,v)_t : 0 \le t < n\} \cup \{\spadesuit\}, \{\spadesuit\}) \quad \text{for } 1 \le i \le w(u,v) \quad (7)$$
$$(\{\heartsuit\}, \{\spadesuit\}, \{\heartsuit\}) \quad (8)$$

The meaning of an element $(u,v)_t$ in a state of $\mathcal{A}$ is that edge (u,v) is the t-th edge (for $0 \le t < n$) of a simple cycle over V, i.e., of a candidate solution for the TSP. The edge weights of the cycle must also appear, in unary notation: if $(u,v)_t$ is part of a state and $w(u,v) = k$, then also $\diamondsuit_{(u,v),1}, \dots, \diamondsuit_{(u,v),k}$ must be part of the state. Hence, a length-n simple cycle $\mathbf{c} = (v_0, \dots, v_{n-1})$ over V is encoded as a state $T(\mathbf{c}) = E(\mathbf{c}) \cup W(\mathbf{c}) \cup \{\heartsuit\}$, where

$$E(\mathbf{c}) = \{(v_0,v_1)_0, (v_1,v_2)_1, \dots, (v_{n-2},v_{n-1})_{n-2}, (v_{n-1},v_0)_{n-1}\}$$

is the set of edges traversed by the cycle, indexed in order of traversal, and

$$W(\mathbf{c}) = \{\diamondsuit_{(v_t,v_{(t+1) \bmod n}),i} : 0 \le t < n, 1 \le i \le w(v_t, v_{(t+1) \bmod n})\}$$

contains the elements encoding the weights of the edges in $E(\mathbf{c})$.

Moreover, consider a state $T \subseteq S$. If $\heartsuit \notin T$, then necessarily $\mathrm{res}_{\mathcal{A}}(T) = \varnothing$, since all reactions have $\heartsuit$ as a reactant. Similarly, $\spadesuit \in T$ implies $\mathrm{res}_{\mathcal{A}}(T) = \varnothing$, since $\spadesuit$ inhibits all reactions. Now suppose $\heartsuit \in T$ and $\spadesuit \notin T$. Reactions (1)–(5) produce $\spadesuit$ from T when any of the following conditions (implying that $T \cap E$ does *not* encode a simple cycle over V) occur:

(1) the state T contains two distinct elements denoting edges occurring as the t-th edge of the candidate cycle;
(2) one of the vertices of the candidate cycle has outdegree greater than 1;
(3) one of the vertices of the candidate cycle has indegree greater than 1;

(4) two consecutive edges do not share an endpoint;
(5) no edge is the t-th edge of the candidate cycle, for some $0 \leq t < n$.

Any set $T \cap E$ where *none* of the above apply encodes a valid simple cycle over V.

Reaction (6) produces ♠ if an edge $(u, v)_t$ occurs, but some element $\diamondsuit_{(u,v),i}$, encoding a unit of the weight of the edge, is missing. Conversely, reaction (7) produces ♠ if a unit of the weight of a missing edge occurs. These reactions are all simultaneously disabled exactly when $T \cap W$ contains the weights of the edges in $T \cap E$. Finally, reaction (8) preserves the ♡. The result function of $\mathcal{A}$ is thus

$$\mathrm{res}_{\mathcal{A}}(T) = \begin{cases} \{\heartsuit\} & \text{if } T \text{ encodes a weighted simple cycle over } V \\ \{\heartsuit, \spadesuit\} & \text{if } \heartsuit \in T,\ \spadesuit \notin T, \text{ but } T \text{ fails to encode} \\ & \quad \text{a weighted simple cycle over } V \\ \varnothing & \text{if } \heartsuit \notin T \text{ or } \spadesuit \in T \end{cases}$$

Hence, the preimages of $\{\heartsuit\}$ are exactly the weighted simple cycles over V. Since each preimage T of $\{\heartsuit\}$ has size $|T| = n + 1 + \sum_{t=0}^{n-1} w(v_t, v_{(t+1) \bmod n})$, a preimage T of $\{\heartsuit\}$ of minimum size corresponds to a shortest tour over V, which can be extracted from T in polynomial time just by listing the elements in $T \cap E$, ordered by their subscript. Since the mapping $(V, w) \mapsto \mathcal{A}$ described by the above construction can be computed in polynomial time, the thesis follows. □

Lemma 5. *For RS* MPP *is metric reducible to the problem of finding, among the possible output strings of a polynomial-time nondeterministic TM, a string having the minimum number of* 1*s.*

Proof. Given any instance $(\mathcal{A} = (S, A), T)$ of the RS minimal preimage problem, let M be the nondeterministic TM which behaves as follows. M guesses a state $U \subseteq S$ and checks whether $\mathrm{res}_{\mathcal{A}}(U) = T$; if this is the case, then M outputs U as a binary string in $\{0,1\}^{|S|}$; M outputs $1^{|S|+1}$, otherwise. Clearly, M works in time $p(n)$ for some polynomial p, and its outputs are all the preimages of T (together with an easily distinguishable dummy output if no preimage exists); in particular, the outputs of M that are minimal with respect to the number of 1s correspond to the smallest preimages of T. □

The following is proved similarly to the equivalence of binary TSP and finding the maximum binary output of a polynomial-time nondeterministic TM [8].

Lemma 6. *Finding a string with minimal number of* 1*s among those output by a nondeterministic polynomial-time TM is metric reducible to the unary TSP.* □

We can now provide lower and upper bounds to the complexity of MPP.

Theorem 7. MPP *for RS is equivalent to the unary TSP under metric reductions. Hence,* MPP $\in \mathbf{FP}^{\mathbf{NP}}$ *and it is* $\mathbf{FP}_{\|}^{\mathbf{NP}}$*-hard under metric reductions.*

Proof. The equivalence is a consequence of Lemmata 4, 5, and 6. Moreover, TSP belongs to $\mathbf{FP}^{\mathbf{NP}}$ which is closed under metric reductions [8]. The travelling salesman with 0/1 weights is hard for $\mathbf{FP}_{\|}^{\mathbf{NP}}$ (see [1]) and can be reduced to the minimal RS preimage problem. □

The complexity of finding the *size* of minimal preimages is given by:

Corollary 8. SMPP *is* $\mathbf{FP}^{\mathbf{NP}[\log n]}$*-complete under metric reductions.*

Proof. The size of the minimal preimage can be found by binary search, using an oracle answering the question "Is there a preimage of size at most k?", which belongs to **NP**. Hence, the problem is in $\mathbf{FP}^{\mathbf{NP}[\log n]}$. The hardness of the problem follows from the fact that the size of a minimal preimage is just the length of the shortest unary travelling salesman tour increased by $n+1$ in the reduction of Lemma 4, and that the unary TSP is $\mathbf{FP}^{\mathbf{NP}[\log n]}$-complete [8]. □

Finally, we can also prove lower and upper bounds to the complexity of finding *the number of* minimal preimages or #MPP in short.

Theorem 9. #MPP *is in* $\#\mathbf{P}^{\mathbf{NP}[\log n]}$ *and it is* $\#\mathbf{P}$*-hard under parsimonious reductions.*

Proof. The following algorithm shows the membership in $\#\mathbf{P}^{\mathbf{NP}[\log n]}$: given $(\mathcal{A}, T)$, compute the size k of the smallest preimage of T by binary search using $\log n$ queries to the oracle (as in the proof of Corollary 8); then, guess a state $U \subseteq S$ with $|U| = k$, and accept if and only if $\mathrm{res}_{\mathcal{A}}(U) = T$. The number of accepting computations corresponds to the number of minimal preimages of T.

In order to prove the $\#\mathbf{P}$-hardness of the problem, we perform a reduction from #SAT (a variant of [5, Theorem 4]). Let $\varphi = \varphi_1 \wedge \cdots \wedge \varphi_m$ be a Boolean formula in conjunctive normal form over the variables $V = \{x_1, \ldots, x_n\}$. Let $\mathcal{A} = (S, A)$ be a RS with $S = V \cup \overline{V} \cup C \cup \{\spadesuit\}$, where $\overline{V} = \{\bar{x}_1, \ldots, \bar{x}_n\}$ and $C = \{\varphi_1, \ldots, \varphi_m\}$, and A consisting of the following reactions:

$$(\{x_i\}, \{\spadesuit\}, \{\varphi_j\}) \qquad \text{for } 1 \le i \le n,\ 1 \le j \le m, \text{ if } x_i \text{ occurs in } \varphi_j \tag{9}$$

$$(\{\bar{x}_i\}, \{\spadesuit\}, \{\varphi_j\}) \qquad \text{for } 1 \le i \le n,\ 1 \le j \le m, \text{ if } \bar{x}_i \text{ occurs in } \varphi_j \tag{10}$$

$$(\{s\}, \{x_i, \bar{x}_i, \spadesuit\}, \{\spadesuit\}) \qquad \text{for } 1 \le i \le n,\ s \notin \{x_i, \bar{x}_i, \spadesuit\} \tag{11}$$

$$(\{x_i, \bar{x}_i\}, \{\spadesuit\}, \{\spadesuit\}) \qquad \text{for } 1 \le i \le n. \tag{12}$$

A state $T \subseteq S$ encodes a valid assignment for φ if, for each $1 \le i \le n$, it contains either x_i or $\bar{x}_i$ (denoting the truth value of variable x_i), but not both, and no further element. The reactions of type (9) (resp., (10)) produce the set of elements representing the clauses satisfied when x_i is assigned a true (resp., false) value. Hence, the formula φ has as many satisfying truth assignments as the number of states $T \subseteq S$ encoding valid assignment such that $\mathrm{res}_{\mathcal{A}}(T) = C$, the whole set of clauses. Any such T contains exactly n elements.

If a state T has n elements or less, but it is not a valid assignment to φ, then there is at least one literal x_i or $\bar{x}_i$ missing in T: thus, either $T = \varnothing$, or reaction (11) is enabled and $\spadesuit \in \mathrm{res}_{\mathcal{A}}(T)$; in both cases $\mathrm{res}_{\mathcal{A}}(T) \neq C$. If T has strictly more than n elements, then either it is an inconsistent assignment, containing both x_i and $\bar{x}_i$ for some i, and in that case $\spadesuit \in \mathrm{res}_{\mathcal{A}}(T) \neq C$ by reaction (12), or it has a subset $T' = T \cap (V \cup \overline{V})$ with $|T'| = n$ such that $\mathrm{res}_{\mathcal{A}}(T') = \mathrm{res}_{\mathcal{A}}(T)$.

Thus, the number of *minimal* preimages of T in $\mathcal{A}$ is exactly the number of assignments satisfying φ, as required. □

4 Ancestor Problems

We now turn our attention to the ancestor problems, which will show a (supposedly) higher complexity. First of all, we need a few technical results, providing us with a **FPSPACE**(poly)-complete function suitable for reductions.

Lemma 10. *The "universal" function $U(M, 1^m, x)$, defined as $M(x)$ if the TM M halts in space m, and undefined otherwise, is* **FPSPACE**(poly)*-complete under many-one reductions.*

Proof. We have $U \in$ **FPSPACE**(poly), since there exist universal TMs having only a polynomial space overhead [8]. Let $R \in$ **FPSPACE**(poly), and let M be a polynomial-space TM for R. Then, there exists a polynomial p bounding both the working space and the output length of M. The mapping $f(x) = (M, 1^{p(|x|)}, x)$ can be computed in polynomial time, and $R(x, U(f(x)))$ holds for all $x \in \operatorname{dom} R$. This proves the **FPSPACE**(poly)-hardness of U. □

Lemma 11. *Let the binary relation $R((M, 1^m, y), x)$ hold if and only if the deterministic TM M, on input x, halts with output y on its tape, $|y| \leq m$, and M does not exceed space m during the computation. Then, the relation R is* **FPSPACE**(poly)*-complete under many-one reductions.*

Proof. The relation R is in **FPSPACE**(poly): a polynomial-space deterministic TM can try all strings x of length at most m, one by one, and simulate M on input x (within space m) until the output y is produced or the strings have been exhausted. We now reduce the function U from Lemma 10 to R. Given an instance $(M, 1^m, x)$ of U, consider the instance $(M', 1^k, 1)$ of R, where

- M' is the TM which on any input y, first simulates M on input x within space k; if $M(x) = y$, then M' outputs 1; otherwise, M' outputs 0.
- $k = p(m)$, where p is the polynomial space overhead needed by M' in order to simulate M.

We have $U(M, 1^m, x) = y$ if and only if M' outputs 1 in space k on input y, that is, if and only if $R((M', 1^k, 1), y)$. Since the mapping $(M, 1^m, x) \mapsto (M', 1^k, 1)$ can be computed in polynomial time, the relation R is **FPSPACE**(poly)-hard. □

By exploiting the ability of RS to simulate polynomial-space TMs [4], we obtain the following result.

Theorem 12. MAP *is complete for* **FPSPACE**(poly) *under metric reductions.*

Proof. The problem is in **FPSPACE**(poly), since a polynomial-space TM can enumerate all states of a RS in order of size, and check whether they lead to the target state (the reachability problem for RS is known to be in **PSPACE** [4]).

In order to prove the **FPSPACE**(poly)-hardness of the problem, we describe a variant of the simulation of polynomial-space TMs by means of RS from [4], where a distinguished state T, encoding the final configuration of the TM, has as minimal ancestors all states encoding the inputs of the TM leading to that

configuration. Given a TM M and a space bound m, let Q and Γ be the set of states and the tape alphabet of M, including a symbol $\sqcup$ for "blank" cells. We define a RS $\mathcal{A}$ having background set

$$S = \{q_i : q \in Q, -1 \leq i \leq m\} \cup \{a_i : a \in \Gamma, 0 \leq i < m\} \cup \{\spadesuit\}.$$

We encode the configurations of M as states of $\mathcal{A}$ as follows: if M is in state q, the tape contains the string $w = w_0 \cdots w_{m-1}$, and the tape head is located on the i-th cell, then the corresponding state of $\mathcal{A}$ is $\{q_i, w_{0,0}, \ldots, w_{m-1,m-1}\}$, with an element q_i representing TM state q and head position i, and m elements corresponding to the symbols on the tape indexed by their position.

A transition $\delta(q, a) = (r, b, d)$ of M is implemented by the following reactions

$$(\{q_i, a_i\}, \{\spadesuit\}, \{r_{i+d}, b_i\}) \qquad \text{for } 0 \leq i < m$$

which update state, head position, and symbol under the tape head. The remaining symbols on the tape (which have an index different from the tape head position) are instead preserved by the following reactions:

$$(\{a_i\}, \{q_i : q \in Q\} \cup \{\spadesuit\}, \{a_i\}) \qquad \text{for } a \in \Gamma,\ 0 \leq i < m$$

If an element encoding TM state and position is not part of the current RS state, the element representing the initial state $s \in Q$ of M, with tape head on the first cell, is produced by the following reactions

$$(\{a_i\}, \{q_j : q \in Q, 0 \leq j < m\} \cup \{\spadesuit\}, \{s_0\}) \qquad \text{for } a \in \Gamma,\ 0 \leq i < m \tag{13}$$

and the simulation of M by $\mathcal{A}$ begins in the next time step with the same tape contents. If M exceeds its space bound, by moving the tape head to the left of position 0 or to the right of position m, the following reactions become enabled

$$(\{q_{-1}\}, \{\spadesuit\}, \{\spadesuit\}) \qquad \text{for } q \in Q$$
$$(\{q_m\}, \{\spadesuit\}, \{\spadesuit\}) \qquad \text{for } q \in Q$$

and produce the universal inhibitor $\spadesuit$, which halts the simulation in the next time step. The universal inhibitor is also produced by the following reactions when the state of $\mathcal{A}$ is not a valid encoding of a configuration of M, namely, when multiple state elements appear:

$$(\{q_i, r_j\}, \{\spadesuit\}, \{\spadesuit\}) \qquad \text{for } q, r \in Q,\ q \neq r,\ 0 \leq i < m,\ 0 \leq j < m$$

or when multiple symbols are located on the same tape cell:

$$(\{a_i, b_i\}, \{\spadesuit\}, \{\spadesuit\}) \qquad \text{for } a, b \in \Gamma,\ a \neq b,\ 0 \leq i < m$$

or when a tape cell does not contain any symbol (recall that a blank cell contains a specified symbol from Γ):

$$(\{s\}, \{a_i : a \in \Gamma\} \cup \{\spadesuit\}, \{\spadesuit\}) \quad \text{for } s \notin \{a_i : a \in \Gamma\} \cup \{\spadesuit\},\ 0 \leq i < m \tag{14}$$

The result function of $\mathcal{A}$ can thus be described as follows:

$$\mathrm{res}_{\mathcal{A}}(T) = \begin{cases} T' & \text{if } T \text{ encodes a configuration of } M \\ & \quad \text{and } T' \text{ its next configuration} \\ T \cup \{s_0\} & \text{if } T \text{ encodes a tape of } M \\ T' \cup \{\spadesuit\} & \text{for some } T' \subseteq S, \text{ if } \spadesuit \notin T \text{ but } T \text{ does} \\ & \quad \text{not encode a configuration of } M \\ \varnothing & \text{if } \spadesuit \in T \text{ or } T = \varnothing \end{cases}$$

Notice that all states of $\mathcal{A}$ encoding configurations of M have $m+1$ elements, and that either $\mathrm{res}_{\mathcal{A}}(T) = \varnothing$ or $\spadesuit \in \mathrm{res}_{\mathcal{A}}(T)$ if $|T| < m$.

Given an instance $(M, 1^m, y)$ of relation R from Lemma 11, we can ask for a minimal state X of $\mathcal{A}$ leading to Y, where Y encodes the configuration of M in its final state, with the string y on the tape.

If there exists a string x such that $M(x) = y$ and the space bound m is never exceeded by M during its computation, then there exists a state $X \subseteq S$ encoding a tape for M containing the input x (padded to length m with blanks) and such that $\mathrm{res}^t_{\mathcal{A}}(X) = Y$ for some $t \geq 0$. We have $|X| = m$, and X is minimal with respect to size among all states leading to Y, since all smaller states lead to $\varnothing$ in at most two steps. Furthermore, from X we can easily recover a string x with $M(x) = y$. Conversely, any state $X \subseteq S$ with $|X| = m$ and $\mathrm{res}^t_{\mathcal{A}}(X) = Y$ necessarily encodes a string x such that $M(x) = y$ within space m.

If $M(x) \neq y$ for all strings x (or all such computations exceed the space bound m), then all states T such that $\mathrm{res}^t_{\mathcal{A}}(T) = Y$, and in particular the minimal ones, contain $m+1$ elements. By observing this fact, we can infer that no input of M produces the output y in space m.

Since the mapping $(M, 1^m, y) \mapsto (\mathcal{A}, Y)$ described by the above construction can be computed in polynomial time, and the answer for R can be extracted from the answer to the minimal RS ancestor search problem in polynomial time, the latter problem is **FPSPACE**(poly)-hard under metric reductions. □

We now deal with the problem of finding the size of a minimal ancestor.

Lemma 13. *Let $f(M, 1^m) = \min\{|x| : \text{the TM } M \text{ accepts } x \text{ in space } m\}$, undefined if no such x exists. Then f is* **FPSPACE**(log)*-complete under many-one reductions.*

Proof. Given $g \in$ **FPSPACE**(log), let G be a deterministic TM computing g in space $p(n)$ for some polynomial p, and let $x \in \Sigma^\star$. Let M be a deterministic TM that, on input y, first simulates $G(x)$, then accepts if and only if $|y| \geq G(x)$. Hence, we have $g(x) = f(M, 1^{q(|x|)})$, where q is a polynomial bound on the space needed by M to simulate G. This proves the hardness of f. The function can be computed in **FPSPACE**(log) by simulating M on all strings of length at most m until one is accepted in space m, then outputting its length. □

Theorem 14. SMAP *is* **FPSPACE**(log)*-complete under metric reductions.*

Proof. The problem is in **FPSPACE**(log), since a polynomial-space TM can find an ancestor U of a state of a RS as in the proof of Theorem 12, outputting the size of U rather than U itself. In order to show the hardness of the problem, we reduce the function f from Lemma 13 to it. Given $(M, 1^m)$, let $\mathcal{A}$ be the RS of Theorem 12, simulating M in space 1^m, and let $\mathcal{A}'$ be $\mathcal{A}$ modified as follows. First of all, we add $\heartsuit$ to S, and we also add it as a reactant to all reactions. Then, the reactions of type (13) are replaced by

$$(\{\heartsuit\}, \{a_i : a \in \Gamma\} \cup \{q_j : q \in Q, -1 \leq j \leq m\} \cup \{\spadesuit\}, \{\sqcup_i\}) \quad \text{for } 0 \leq i < m \tag{15}$$

$$(\{\heartsuit\}, \{q_j : q \in Q, -1 \leq j \leq m\} \cup \{\spadesuit\}, \{s_0\}) \tag{16}$$

The reactions of type (15) complete the tape of the TM by producing a blank symbol in position i if no symbol a_i and no state q_j occur. Reaction (16) produces the initial state of M in position 0 when no other state element occurs.

The reactions of type (14) are replaced by

$$(\{q_j\}, \{a_i : a \in \Gamma\} \cup \{\spadesuit\}, \{\spadesuit\}) \quad \text{for } q \in Q,\ 0 \leq i < m,\ -1 \leq j \leq m \tag{17}$$

which give an error (producing $\spadesuit$) when a tape symbol is missing, but only if a state element is already present. Finally, we add the reaction $(\{\heartsuit\}, \{\spadesuit\}, \{\heartsuit\})$, which preserves the $\heartsuit$ element.

The behaviour of $\mathcal{A}'$ differs from $\mathcal{A}$ in the following ways. The input string $x = x_0 \cdots x_{n-1}$ of M is provided as a state $X = \{x_{0,0}, \ldots, x_{n-1,n-1}, \heartsuit\}$. In the first step of $\mathcal{A}'$, the tape is completed by adding blanks and the initial state of M (reactions (15)–(16)); this produces the state $X \cup \{\sqcup_n, \ldots, \sqcup_{m-1}, s_0\}$, which encodes the initial configuration of M. The simulation of M then proceeds as for $\mathcal{A}$ (with the additional element $\heartsuit$ always present).

The ancestors of state T describing an accepting configuration of M (empty tape, accepting state, head in position 0) and of minimal size encode the initial input x of M together with $\heartsuit$, if M accepts at least one string in space m (hence, such ancestors have size at most $m + 1$); if no string is accepted, the minimal ancestors of T all have size at least $m + 2$, by the same reasoning as in the proof of Theorem 12. Hence, $f(M, 1^m) + 1$ is the size of a minimal ancestor of T, if the latter is at most $m + 1$, and $f(M, 1^m)$ is undefined otherwise: this defines a metric reduction of f to this problem. □

Remark 15. The problem of Theorem 14 is actually complete under metric reductions that only increase linearly the length of the output; the class **FPSPACE**(log) is closed under such reductions, but not under general metric reductions.

Finally, we show that counting the number of minimal ancestors has the same complexity as finding one of them.

Lemma 16. *Given a TM M, a unary integer 1^m, and a string y, computing the number of strings x of length at most m such that $M(x) = y$ in space m is* **FPSPACE**(poly)*-complete under many-one reductions.*

Proof. Recalling that **FPSPACE**(poly) = ♮**PSPACE**, the following nondeterministic polynomial-space TM has the required number of accepting computations: on input $(M, 1^m, y)$, it guesses a string x of length at most m (this requires polynomially many guesses); then it simulates M on x, accepting if M outputs y without exceeding space m, and rejecting otherwise.

Given $f \in$ ♮**PSPACE**, let N be a nondeterminstic, polynomial-space TM with $f(x)$ accepting computations on input x; let $p(n)$ be both a space bound for N and bound on the number of nondeterministic choices it makes. Consider the following polynomial-space TM M: on input $z \in \{0,1\}^\star$, it simulates a computation of N on input x, but replaces the nondeterministic choices of N with deterministic lookups to successive bits of z. If N exceeds space $m = p(|x|)$, or halts without having made exactly $|z|$ nondeterministic choices, or the simulated computation of N rejects, then M writes 0 as output; otherwise, M outputs 1.

Hence, M outputs 1 once for each accepting computation of N, that is, for exactly $f(x)$ input strings. Since the mapping $x \mapsto (M, 1^{p(|x|)}, 1)$ can be computed in polynomial time, the **FPSPACE**(poly)-hardness of the problem follows. □

Theorem 17. #MAP *is* **FPSPACE**(poly)*-complete under metric reductions.*

Proof. The problem is in **FPSPACE**(poly), since a polynomial-space TM can compute the size of a minimal ancestor of a state T of a RS, then enumerate all states of the same size and count how many of them lead to T.

Let $(M, 1^m, y)$ be an instance of the problem of Lemma 16, and let $\mathcal{A}$ be the RS simulating M as in the proof of Theorem 12. Let $Y = \{f_0, z_{0,0}, \ldots, z_{m-1,m-1}\}$ be the state of $\mathcal{A}$ encoding the final configuration of M with output y, where $z = y_0 \cdots y_{k-1} \sqcup^{m-k}$ is y padded to length m with blanks and $f \in Q$ is the final state of M. In order to distinguish the presence or absence of at least a string x such that $M(x) = y$, we add a large number of ancestors of Y having size $m+1$, ensuring that are minimal only if no such string exists. Let $\mathcal{A}'$ be $\mathcal{A}$ augmented with the following reactions:

$$(\{a_i, ♠\}, \{q_j : q \in Q, -1 \leq j \leq m\}, \{z_{i,i}, f_0\}) \quad \text{for } a \in \Gamma,\ 0 \leq i < m. \tag{18}$$

When ♠ is present, and all q_j are missing, these reactions map each element representing a symbol in tape cell i, to the symbol z_i in tape cell i, together with the final state f of M in position 0. In particular, when at least one symbol per position i is present, the whole target state Y is produced. Hence, these reactions introduce exactly $|\Gamma|^m$ new ancestors of Y of size $m+1$. The state Y then has at least $|\Gamma|^m + 1$ ancestors of size $m+1$, including Y itself. From the proof of Theorem 12 we may infer that the maximum number of ancestors of Y of size m is $|\Gamma|^m$. The number of strings x of length at most m such that $M(x) = y$ in space m is then equivalent to the number of minimal ancestors of Y for $\mathcal{A}'$, if and only if this number is at most $|\Gamma|^m$. If the number is larger than $|\Gamma|^m$, then the minimal ancestors have size $m+1$, indicating that no such string x exists. This defines a metric reduction, proving the **FPSPACE**(poly)-hardness of #MAP. □

5 Conclusions

We investigated the problem of finding the minimal preimage of a state of a RS and proved that this problem is equivalent, under metric reductions, to finding a minimal TSP tour when the weights are expressed in unary. We also studied the complexity of finding a minimal ancestor of a given state and showed that it is as hard as simulating a polynomial-space TM (with polynomial-length output). Furthermore, we have investigated the complexity of other problems related to preimages (resp., ancestors): finding the size and the number of minimal preimages (resp., ancestors). All these problems were proved to be intractable.

In the future we plan to continue the exploration of problems related to preimages and ancestors of RS. In the more general model [3], RSs behave as interactive processes, where new entities are introduced at every time step by means of a context sequence. Under which conditions does the presence of a context sequence increase the complexity of the problems we considered? We are also interested in questions related to the approximability of the aforementioned problems and the complexity of finding a minimal ancestor that is not "too far" from the target state.

References

1. Chen, Z.Z., Toda, S.: On the complexity of computing optimal solutions. International Journal of Foundations of Computer Science 2(3), 207–220 (1991)
2. Ehrenfeucht, A., Rozenberg, G.: Basic notions of reaction systems. In: Calude, C.S., Calude, E., Dinneen, M.J. (eds.) Developments in Language Theory, 8th International Conference, DLT 2004, Lecture Notes in Computer Science, vol. 3340, pp. 27–29. Springer (2005)
3. Ehrenfeucht, A., Rozenberg, G.: Reaction systems. Fundamenta Informaticae 75, 263–280 (2007)
4. Formenti, E., Manzoni, L., Porreca, A.E.: Cycles and global attractors of reaction systems. In: Jürgensen, H., Karhumäki, J., Okhotin, A. (eds.) Descriptional Complexity of Formal Systems, 16th International Workshop, DCFS 2014. Lecture Notes in Computer Science, vol. 8614, pp. 114–125. Springer (2014)
5. Formenti, E., Manzoni, L., Porreca, A.E.: Fixed points and attractors of reaction systems. In: Beckmann, A., Csuhaj-Varjú, E., Meer, K. (eds.) Language, Life, Limits, 10th Conference on Computability in Europe, CiE 2014, Lecture Notes in Computer Science, vol. 8493, pp. 194–203. Springer (2014)
6. Krentel, M.W.: The complexity of optimization problems. Journal of Computer and System Sciences 36, 490–509 (1988)
7. Ladner, R.E.: Polynomial space counting problems. SIAM Journal on Computing 18(6), 1087–1097 (1989)
8. Papadimitriou, C.H.: Computational Complexity. Addison-Wesley (1993)
9. Porreca, A.E., Murphy, N., Pérez-Jiménez, M.J.: An optimal frontier of the efficiency of tissue P systems with cell division. In: García-Quismondo, M., Macías-Ramos, L.F., Păun, Gh., Valencia-Cabrera, L. (eds.) Tenth Brainstorming Week on Membrane Computing, vol. II, pp. 141–166. Fénix Editora (2012)
10. Salomaa, A.: Minimal and almost minimal reaction systems. Natural Computing 12(3), 369–376 (2013)